

Pierre Bailly
Christine Carrère

Statistiques descriptives
Exercices

Collection « Libres Cours Économie »

Presses universitaires de Grenoble
BP 47 – 38040 Grenoble cedex 9
Tél. : 04 76 82 56 52 – pug@pug.fr / www.pug.fr

Chapitre 1

Les outils

Nous présentons dans ce chapitre des exercices sur quatre thèmes : les nomenclatures et les types de variable, les tableaux statistiques, les représentations graphiques, l'utilisation des pourcentages et des taux.

1. TYPES DE VARIABLE, VARIABLE OU CARACTÈRE

Mots-clés

Variable discrète, variable continue, caractère qualitatif

Énoncé

Quelle est la nature des caractères ci-dessous ?

Nombre d'actions vendues chaque jour à la bourse

Rémunérations des enseignants d'un lycée

Indicateur du moral des ménages

Écart de rémunération entre hommes et femmes

Les pays de l'Union européenne

Les niveaux de formation des salariés

Les formes de contrat de travail

Taux de croissance du PIB

Prix à la consommation

Solde commercial

Nombre de personnes par ménages

Corrigé

Nombre d'actions vendues chaque jour à la bourse	variable discrète
Rémunérations des enseignants d'un lycée	variable quantitative continue
Indicateur du moral des ménages	variable qualitative ordonnée
Écart de rémunération entre hommes et femmes	variable continue

Les pays de l'Union européenne	caractère qualitatif
Les niveaux de formation des salariés	variable ordonnée
Les formes de contrat de travail	caractère qualitatif
Taux de croissance du PIB	variable quantitative
Prix à la consommation	variable quantitative
Solde commercial	variable quantitative
Nombre de personnes par ménage	variable statistique discrète.

2. UTILISATION DE LA NAF

Mots-clefs

Caractère qualitatif, nomenclature

Énoncé

1. Quel est le code de la nomenclature NAF 31 correspondant à la « Fabrication d'équipements électriques et électroniques » ?
2. Quelle est l'activité codée LL ?

Vous trouverez la nomenclature nécessaire p. 125 du livre de cours.

Corrigé

1. Le code de la nomenclature NAF correspondant à la « Fabrication d'équipements électriques et électroniques » est obtenu par lecture de la nomenclature. Le code est DL.
2. Le code LL correspond à l'activité « Administration publique ».

3. UTILISATION D'UNE NOMENCLATURE DE L'UNION EUROPÉENNE

Mots-clefs

Caractère qualitatif

Énoncé

Quelle est l'activité correspondant au E de la NACE-CE ?

Vous trouverez la nomenclature nécessaire p. 126 du livre de cours.

Corrigé

L'activité correspondant au code E est la « Production et distribution d'électricité, de gaz et d'eau ».

4. NIVEAUX DE FORMATION

Mots-clefs

Caractère qualitatif ordonné

Énoncé

Quel est le niveau de formation d'un étudiant qui, ayant suivi les cours de première année du DEUG de sociologie, n'a pas obtenu son passage en seconde année et quitte l'université ?

Vous trouverez la nomenclature nécessaire page 126 du livre de cours.

Corrigé

Le niveau de formation de cet étudiant sera IV, plus précisément IV sup. La nomenclature des formations est un caractère qualitatif ordonné.

5. NOMBRE DE PERSONNES DANS LES MÉNAGES

Mots-clefs

Variable discrète, tableau statistique

Énoncé

Nous disposons de la distribution des ménages selon leur composition.

Ménages suivant le nombre de personnes du ménage en France en 1995

Ensemble (en milliers)	23 126
Soit en pourcentage, suivant le nombre de personnes dans le ménage	
1 personne	29,2
2 personnes	31,8
3 personnes	16,8
4 personnes	14,2
5 personnes et plus	8,0

Source : TEF 1998/99

1. Construisez le tableau statistique en calculant les effectifs pour chacune des catégories de ménages.
2. Combien de personnes ont été comptées dans cette étude ?

Corrigé

Un ménage est constitué des personnes occupant une même unité d'habitation.

1. La construction du tableau statistique nécessite de calculer l’effectif de chaque catégorie de ménages. Il est obtenu en multipliant le nombre total des ménages par son importance relative. Les résultats ont été arrondis au millier d’unité près.

Par exemple, le calcul du nombre de ménages comprenant trois personnes est le suivant :

Effectifs des ménages de trois personnes = $23\,126 \cdot 0,168 = 3\,885,168 \cong 3\,885$

Effectifs des ménages suivant le nombre de personnes dans le ménage

Valeurs de la variable	Fréquences	Pourcentages	Effectifs (milliers)
x_i	f_i	p_i	n_i
1 personne	0,292	29,2	6 753
2 personnes	0,318	31,8	7 354
3 personnes	0,168	16,8	3 885
4 personnes	0,142	14,2	3 284
5 personnes et plus	0,080	8,0	1 850
Total	1,000	100,0	23 126

2. Pour calculer le nombre de personnes concernées par l’étude, nous devons faire une hypothèse sur la taille des ménages de la classe « 5 personnes et plus ». Dans le cas où nous considérerions que la taille moyenne de cette catégorie de ménages est de 6, nous obtiendrions une population de 57 352 milliers de personnes. Si nous retenons l’hypothèse de 7 personnes par ménage dans cette classe, nous avons une population de 59 202 milliers de personnes.

Si nous retenons la première hypothèse, le détail du calcul est :

$1 \cdot 6\,753 + 2 \cdot 7\,354 + 3 \cdot 3\,885 + 4 \cdot 3\,284 + 6 \cdot 1\,850 = 57\,352$

Si nous retenons la seconde hypothèse, le détail du calcul est :

$1 \cdot 6\,753 + 2 \cdot 7\,354 + 3 \cdot 3\,885 + 4 \cdot 3\,284 + 7 \cdot 1\,850 = 59\,202.$

6. APPELS TÉLÉPHONIQUES

Mots-clefs

Tableau statistique, variable continue

Énoncé

Vous disposez d’une facture détaillée des appels d’un abonné à France Télécom sur la période 09/03/05-05/05/05.

Effectuez le regroupement en classes de ces données selon les trois variables continues. Vous veillerez à ce que les classes que vous avez choisies respectent les conditions qu'on attend d'un regroupement en classes.

1. Plage horaire de l'appel
2. Durée des appels
3. Montant des appels.

Détail des appels

Date	Heure	Durée	Montant
jj.mm		mm:ss	Hors taxe en centime d'euro
09.03	11:12	11:25	2,570
11.03	21:16	06:38	1,040
14.03	09:40	01:29	3,070
15.03	17:04	00:19	0,610
15.03	17:28	02:17	1,970
15.03	18:31	02:24	2,070
15.03	20:27	07:10	1,100
15.03	20:53	06:47	3,160
16.03	16:15	14:38	3,310
16.03	20:05	04:27	2,190
17.03	10:41	05:25	1,180
17.03	14:36	00:34	2,460
17.03	15:17	10:21	2,320
17.03	16:06	07:39	1,690
17.03	21:05	12:17	1,690
23.03	11:31	00:35	2,460
24.03	13:30	21:14	4,850
24.03	21:22	27:02	3,400
25.03	16:34	00:10	2,460
25.03	21:19	00:37	2,460
26.03	18:58	01:44	4,280
27.03	16:31	03:44	0,700
27.03	22:51	03:47	0,710
30.03	15:34	00:37	2,460
01.04	18:48	05:45	1,250
01.04	21:55	10:52	1,530
03.04	10:09	00:44	3,070
04.04	20:55	03:18	0,650
05.04	20:47	06:35	3,080
06.04	20:56	08:00	1,190
07.04	20:20	07:41	1,160
10.04	17:14	00:42	1,840
21.04	08:58	05:23	1,170
21.04	09:28	04:04	0,860
21.04	09:34	06:35	1,450

Date	Heure	Durée	Montant
jj.mm		mm:ss	Hors taxe en centime d'euro
24.04	14:21	00:26	2,460
26.04	09:05	00:31	3,070
26.04	20:50	14:01	1,890
26.04	21:40	03:33	0,680
26.04	21:44	10:29	1,480
27.04	20:42	03:03	0,620
27.04	20:48	08:19	1,230
27.04	20:57	08:49	1,290
27.04	21:11	04:49	0,830
27.04	21:17	05:10	0,870
27.04	21:22	03:16	0,650
29.04	20:29	05:46	0,940
01.05	14:01	01:33	3,070
01.05	17:11	03:17	0,650
03.05	20:38	09:40	1,390
03.05	20:57	12:22	1,700
04.05	16:13	20:31	4,680
04.05	20:28	03:04	0,620
05.05	14:19	08:36	1,910
05.05	14:49	26:28	6,060
05.05	19:54	05:48	14,400

Corrigé

1. Pour effectuer le dépouillement sur la plage horaire de l'appel, nous constatons qu'aucun appel n'a lieu avant 8h et aucun après 23h. Nous choisissons des plages de deux heures démarrant à 8h, la dernière plage sera une plage d'une heure.

Plage horaire de l'appel

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[8 ; 10[5			
[10 ; 12[4			
[12 ; 14[1			
[14 ; 16[7			
[16 ; 18[9			
[18 ; 20[4			
[20 ; 22[25			
[22 ; 23]		1			
Total		56			

Avant d'aller plus loin dans le tableau, nous constatons un déséquilibre au niveau de la plage entre 20h et 22h. La perte d'information est importante et peut facilement être réduite en utilisant des plages d'une heure pour ce créneau horaire.

Plage horaire de l'appel

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[8 ; 10[9	5	0,089	8,9	0,089
[10 ; 12[11	4	0,071	7,1	0,161
[12 ; 14[13	1	0,018	1,8	0,179
[14 ; 16[15	7	0,125	12,5	0,304
[16 ; 18[17	9	0,161	16,1	0,464
[18 ; 20[19	4	0,071	7,1	0,536
[20 ; 21[20,5	15	0,268	26,8	0,804
[21 ; 22]	21,5	10	0,179	17,9	0,982
[22 ; 23]	22,5	1	0,018	1,8	1,000
Total		56	1,000	100,0	

2.

Durée des appels

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 5[26			
[5 ; 10[16			
[10 ; 15[10			
[15 ; 20[0			
[20 ; 25[2			
[25 ; 30]		2			
Total		56			

Avant d'aller plus loin, constatons que les premières classes sont disproportionnées par rapport aux autres et qu'il y a une classe vide. Un autre découpage possible est explicité dans le tableau suivant.

Durée des appels (regroupement plus judicieux)

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 2[1	13	0,232	23,2	0,232
[2 ; 4[3	10	0,179	17,9	0,411
[4 ; 6[5	9	0,161	16,1	0,571
[6 ; 8[7	7	0,125	12,5	0,696
[8 ; 10[9	5	0,089	8,9	0,786
[10 ; 12[11	4	0,071	7,1	0,857
[12 ; 14[13	2	0,036	3,6	0,893
[14 ; 16[15	2	0,036	3,6	0,929
[16 ; 30[23	4	0,071	7,1	1,000
Total		56	1,000	100,0	

3.

Montants des appels

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 1[0,5	13	0,232	23,2	0,232
[1 ; 2[1,5	20	0,357	35,7	0,589
[2 ; 3[2,5	10	0,179	17,9	0,768
[3 ; 4[3,5	8	0,143	14,3	0,911
[4 ; 5[4,5	3	0,054	5,4	0,964
[5 ; 10[7,5	1	0,018	1,8	0,982
[10 ; 15[12,5	1	0,018	1,8	1,000
Total		56	1	100,0	

7. TABLE DE MORTALITÉ

Mots-clefs

Variable continue, tableau statistique

Énoncé

Extrait de la table de mortalité de la génération féminine française de 1899

Age exact	Survivants à l'âge exact
0	100 000
1	84 883
2	82 247
3	80 843
4	79 995
5	79 186
6	78 763
7	78 411

Source : « La mortalité par génération en France depuis 1899 »,
Travaux et documents, Cahier INED n°63, 1973

1. Présentez le tableau statistique de la variable « âge du décès » sous sa forme habituelle.
2. Donnez la signification concrète de chacune des colonnes du tableau statistique obtenu.

Corrigé

1. L'étude porte sur 100 000 filles nées en 1899 dont le décès est survenu avant l'âge de 7 ans.

Comment obtient-on les effectifs du tableau ? Nous allons prendre comme exemple le cas de la première classe. Nous savons qu'il y a eu 100 000 naissances ; un an plus tard seules 84 883 femmes sont encore vivantes, le nombre de décès est donc de $100\,000 - 84\,883 = 15\,117$. Nous répétons le raisonnement pour tous les âges, ce qui nous permet de construire le tableau statistique.

Répartition de la génération féminine française de 1899 suivant l'âge du décès

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 1[0,5	15117	0,700	70,0	0,700
[1 ; 2[1,5	2636	0,122	12,2	0,822
[2 ; 3[2,5	1404	0,065	6,5	0,887
[3 ; 4[3,5	848	0,039	3,9	0,927
[4 ; 5[4,5	809	0,037	3,7	0,964
[5 ; 6[5,5	423	0,020	2,0	0,984
[6 ; 7]	6,5	352	0,016	1,6	1,000
		21589	1,000	100,0	

2. La première colonne reprend la tranche d'âge des décès. La deuxième représente l'âge moyen du décès par classe annuelle. La colonne des effectifs donne le nombre de femmes décédées dans la tranche d'âge considérée. Au total 21589 filles sont décédées avant l'âge de 7 ans. La colonne suivante donne la fréquence des femmes décédées dans la tranche d'âge, ainsi que la troisième qui exprime la même chose en pourcentage. La dernière colonne donne la fréquence des femmes décédées avant la borne supérieure de la tranche d'âge. La fréquence signifie que 92,7 % des femmes mortes avant l'âge de 7 ans sont mortes avant l'âge de 4 ans. Le tableau montre que la mortalité féminine est très forte au cours de la première année (70 % des décès).

8. BILAN DES APPORTS ET DES USAGES DE L'EAU

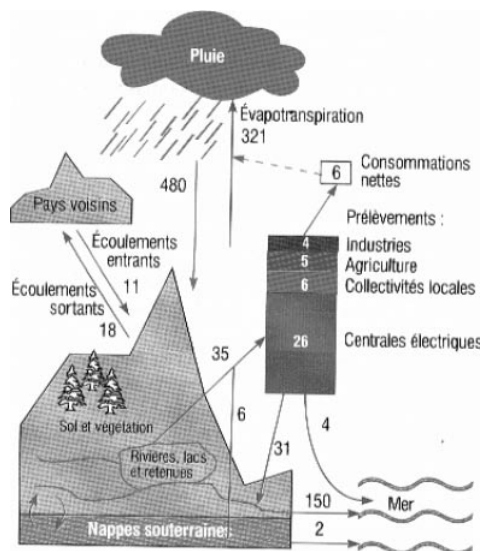
Mots-clefs

Analyse de graphique

Énoncé

L'exercice consiste à commenter un pictogramme

Bilan des apports et des usages des eaux continentales (en milliards de m³/an)



Source : TEF 1998/1999, Paris : INSEE

1. Quelle est la quantité nette d'eau consommée par les centrales électriques ?
2. Quel est le bilan des échanges d'eau avec les pays voisins ?
3. Quelle est l'équation d'équilibre des usages humains de l'eau ?
4. Quelle est l'équation d'équilibre des eaux continentales ?

Corrigé

1. Pour calculer la part des centrales électriques dans la quantité nette d'eau consommée, nous supposons que la « consommation » nette, en fait l'évaporation, est proportionnelle à la quantité utilisée. Cette hypothèse est sans doute une sous-estimation dans le cas des centrales électriques. Pour un usage total de 41 milliards de m³ par an, 35 provenant des précipitations et 6 des nappes phréatiques, nous obtenons une consommation de :

$$6 \cdot \frac{26}{41} = 3,8 \text{ milliards de m}^3$$

2. La France reçoit 11 milliards de m³ par écoulement, les écoulements vers les pays voisins s'élèvent à 18 milliards de m³, donc le solde des échanges extérieurs est de - 7 milliards de m³.
3. Les activités humaines prélèvent 35 milliards de m³ dans « les rivières, lacs et retenues » et 6 milliards de m³ dans les nappes phréatiques soit 41 milliards de m³.

Les usages humains se traduisent par une évapotranspiration de 6 milliards de m³, un écoulement de 4 milliards de m³ vers la mer et de 31 milliards de m³ dans les « rivières, lacs et retenues », soit également 41 milliards de m³.

Emplois – ressources des usages humains (milliards de m³)

Emplois		Ressources	
Consommations nettes (évaporation)	6	Prélèvements dans les « rivières, lacs et retenues »	35
Écoulement vers la mer	4	Ponctions dans les nappes phréatiques	6
Rejets vers les « rivières, lacs et retenues »	31		
Total	41	Total	41

4. Les apports sont de 480 milliards de m³ de précipitations (pluie, neige) moins 7 milliards de m³ provenant des échanges avec les pays voisins. Les apports sont de 473 milliards de m³.

Les usages comprennent 321 milliards de m³ sous forme d'évapotranspiration, 150 milliards de m³ sous forme de ruissellement et 2 milliards de m³ d'écoulement des nappes phréatiques vers la mer. Nous obtenons bien 473 milliards de m³.

Emplois – ressources des eaux continentales (milliards de m³)

Emplois		Ressources	
Évapotranspiration	321	Précipitations	480
Ruissellement	150	Échanges avec les pays voisins	– 7
Écoulement vers la mer	2		
Total	473	Total	473

9. ANALYSE DE LA MÉDICALISATION

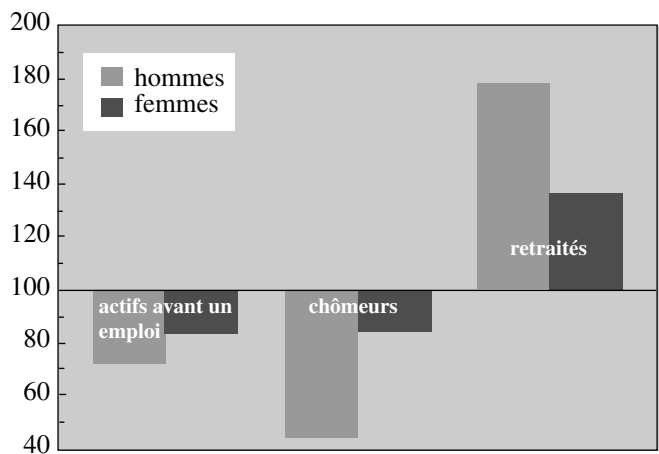
Mots-clefs

Commentaire de graphique, indice

Énoncé

Commentez le graphique ci-dessous selon les catégories et selon les genres.

Disparité de la médicalisation



Source : TEF 98/99

Note : 100 correspond à l'indice du nombre moyen de séances de médecins par sexe soit 5,28 séances pour les hommes et 8,36 séances pour les femmes.

Corrigé

La disparité de la médicalisation par catégorie apparaît dans ce graphique. Les actifs et les chômeurs ont moins recours au médecin que les retraités comme on pouvait s'y attendre. Les actifs et les chômeurs ont globalement une consommation inférieure à la moyenne.

Globalement, les femmes consultent beaucoup plus souvent que les hommes soit 5,28 séances pour les hommes et 8,36 séances pour les femmes dans un rapport de $\frac{8,36}{5,28} = 1,583 \cong 1,6$. Cette situation s'inverse pour la catégorie « retraités ».

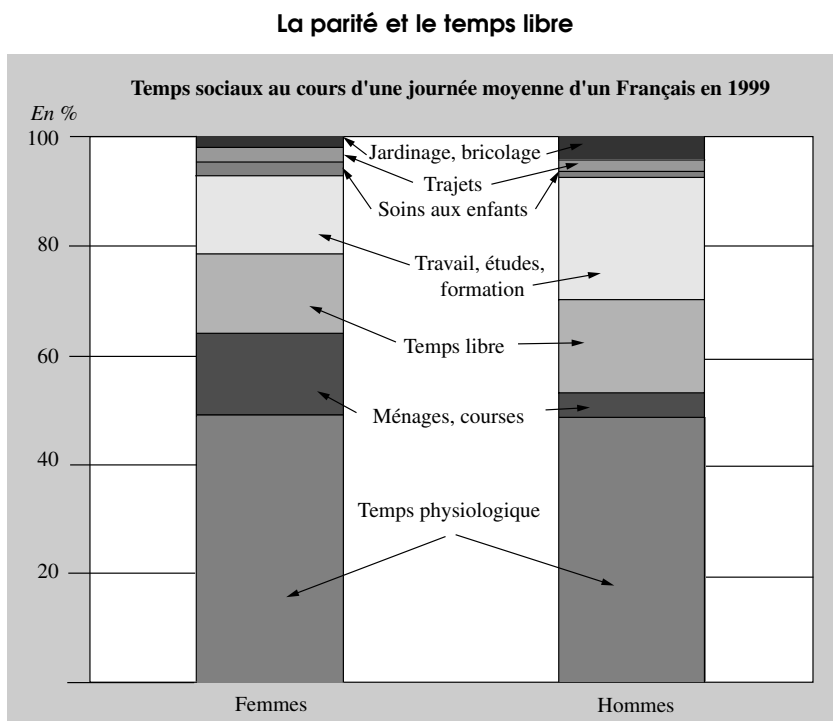
Les femmes « actives » ont un recours au médecin aussi fréquent que les « chômeuses ». Elles ont une consommation médicale plus importante que les hommes de même catégorie. Pour la population active (actifs ayant un emploi et chômeurs), les hommes s'occupent moins de leur santé que les femmes. L'écart entre les hommes et les femmes est beaucoup plus important dans la catégorie « chômeurs ». Les hommes au chômage réduisent très sensiblement leur consommation médicale auprès des médecins alors que les femmes en profitent « pour mieux se soigner ». Par contre les hommes retraités ont une consommation médicale bien supérieure à celle des femmes. Faut-il y voir les effets d'une moindre consommation antérieure ?

10. STRUCTURE DES TEMPS SOCIAUX FÉMININS ET MASCULINS

Mots-clefs

Diagramme à cumul interne, analyse de diagramme

Énoncé



Source : TEF 98/99

Quelles informations vous donnent ce graphique sur les différences dans les temps sociaux entre les hommes et les femmes ?

Corrigé

Les temps sociaux se subdivisent en quatre périodes : le temps « physiologique » consacré à dormir, se laver, manger, etc., le temps de travail professionnel ou d'études, celui consacré aux travaux domestiques tels que le ménage, la lessive, les courses, etc., et le temps des loisirs qui comprend les promenades, la télévision, la pratique d'un sport, la lecture, etc.

Les temps physiologiques sont proches pour les hommes et les femmes avec un très léger avantage aux femmes.

Les temps de « travail, études, formation » sont dans un rapport du simple au double entre les hommes et les femmes. Cela tient d'une part au fait que les femmes inactives sont plus nombreuses que les hommes inactifs et d'autre part que les femmes ont plus souvent que les hommes des emplois à temps partiel. Les emplois à temps partiel sont en majorité occupés par les femmes. Cette dernière information provient d'une connaissance du domaine de l'économie du travail ce qui permet une analyse plus complète du graphique.

Le travail domestique est essentiellement un temps féminin en particulier pour ce qui concerne le ménage, les courses et les soins aux enfants. Le temps féminin est plus que le double du temps masculin dans ces domaines. L'activité domestique est une activité féminine.

En définitive, les femmes disposent, en moyenne de légèrement moins de temps libre que les hommes, la différence reste néanmoins peu significative.

D'un point de vue global, ce graphique donne l'impression qu'entre les hommes et les femmes, il y a échange entre temps de travail (masculin) et activité domestique (féminine). Le modèle traditionnel de partage des tâches entre les hommes et les femmes subsiste pour l'essentiel. Il faudrait, pour préciser les évolutions, disposer de données pour des périodes antérieures.

11. APPELS TÉLÉPHONIQUES

Mots-clefs

Variable discrète, diagramme en bâtons, diagramme cumulatif

Énoncé

Fréquences des appels téléphoniques

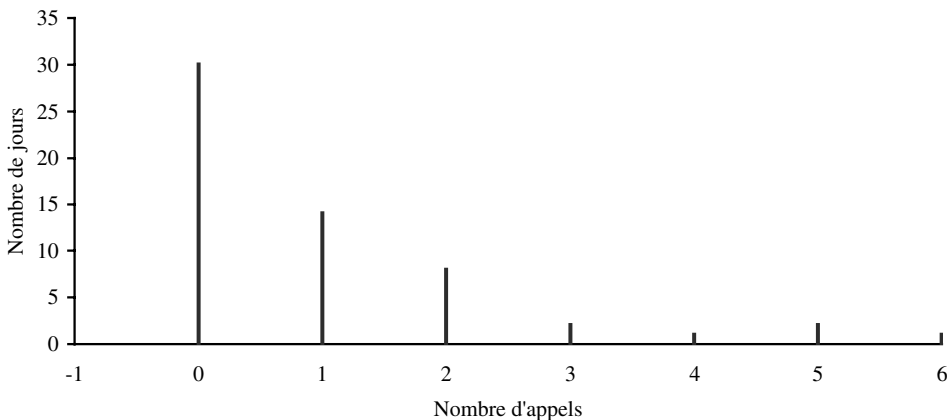
Nombre d'appels	Nombre de jours	Fréquences	Fréquences cumulées
x_i	n_i	f_i	F_i
0	30	0,517	0,517
1	14	0,241	0,759
2	8	0,138	0,897
3	2	0,034	0,931
4	1	0,017	0,948
5	2	0,034	0,983
6	1	0,017	1,000
	58	1,000	

1. Construisez le graphique des effectifs des appels.
2. Construisez le diagramme cumulatif.

Corrigé

1. Nous construisons un diagramme en bâtons appelé diagramme différentiel car il représente les différentes modalités de la variable. Ce diagramme s'impose puisque nous avons une variable discrète.

Nombre d'appels par jour

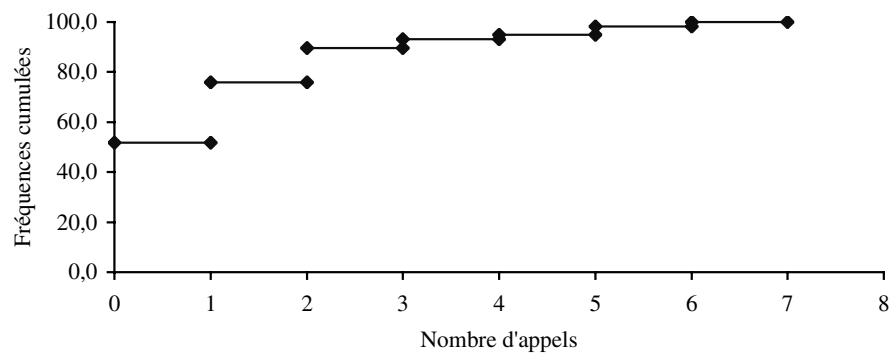


2. Le diagramme cumulatif est également appelé diagramme intégral au sens de l'intégration mathématique. Il représente le graphe des fréquences cumulées. C'est un graphique « en escalier », les valeurs ne sont connues que sur des intervalles en raison même de la nature de la variable.

Tableau des appels cumulés

Nombre d'appels	Nombre de jours	Fréquences cumulées
x_i	n_i	F_i
0	30	0,517
1	14	0,759
2	8	0,897
3	2	0,931
4	1	0,948
5	2	0,983
6	1	1,000
	58	

Diagramme cumulatif



12. ÉVOLUTIONS DES PROFESSIONS ET CATÉGORIES SOCIALES

Mots-clefs

Représentation en secteurs, caractère qualitatif, représentation en barre

Énoncé

Les tableaux suivants fournissent des informations sur l'importance des profes-
sions et catégories sociales dans la population active occupée en 1982, en 1990
et en 2005.

1. Définissez en quelques lignes ces différentes catégories.
2. Donnez des représentations graphiques qui fassent apparaître l'importance relative des différentes catégories sociales pour l'année 1982, pour l'année 1990 et pour l'année 2005.

3. Donnez une représentation graphique qui fasse apparaître les évolutions entre les différentes années.

Les professions et catégories sociales

Catégories	1982	1990	2005
Agriculteurs exploitants	1 475	1 162	651
Artisans, commerçants, chefs d'entreprises	1 835	1 820	1 505
Cadres et professions intellectuelles supérieures	1 895	2 435	3 660
Professions intermédiaires	3 971	4 308	5 745
Employés	6 247	5 925	7 232
Ouvriers	7 749	6 265	5 972

Source : INSEE

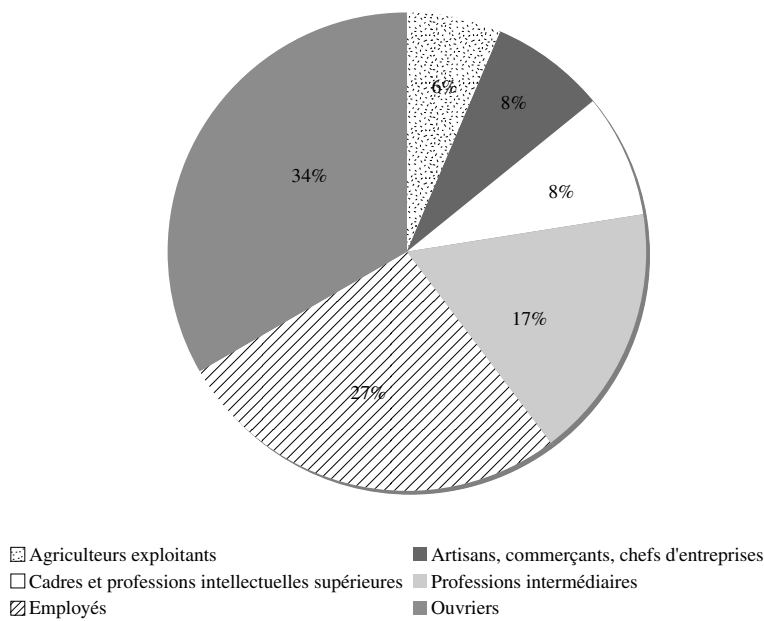
Corrigé

1. Voir correction de l'exercice 25 page 56.
2. Nous utilisons une représentation en secteur qui fait bien apparaître l'importance relative de chaque catégorie. Le tableau ci-dessous fournit les données pour la construction des graphiques.

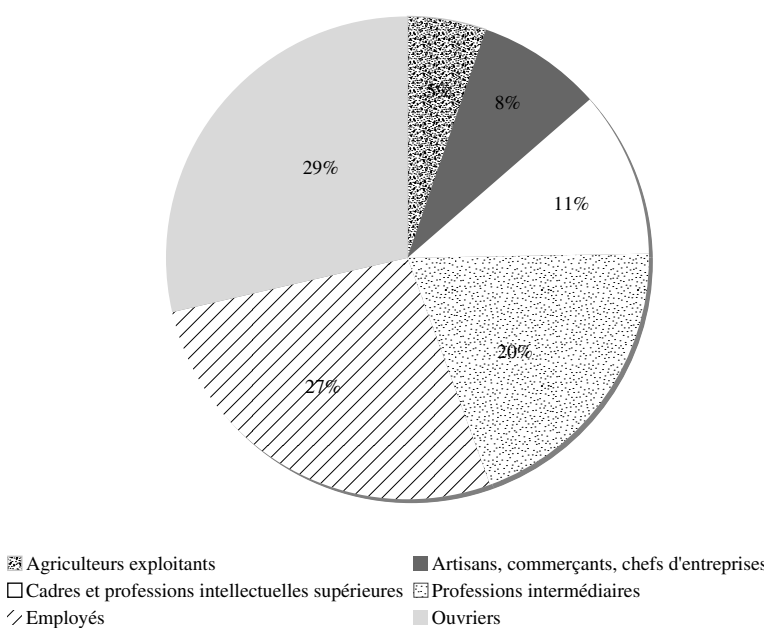
Importance relative des différentes catégories

Catégorie	Effectifs			Fréquences		
	n_i (milliers)			f_i (%)		
	1982	1990	2005	1982	1990	2005
Agriculteurs exploitants	1 475	1 162	651	6,4	5,3	2,6
Artisans, commerçants, chefs d'entreprises	1 835	1 820	1 505	7,9	8,3	6,1
Cadres et professions intellectuelles supérieures	1 895	2 435	3 660	8,2	11,1	14,8
Professions intermédiaires	3 971	4 308	5 745	17,1	19,7	23,2
Employés	6 247	5 925	7 232	27,0	27,0	29,2
Ouvriers	7 749	6 265	5 972	33,4	28,6	24,1
Ensemble	23 172	21 915	24 765	100,0	100,0	100,0

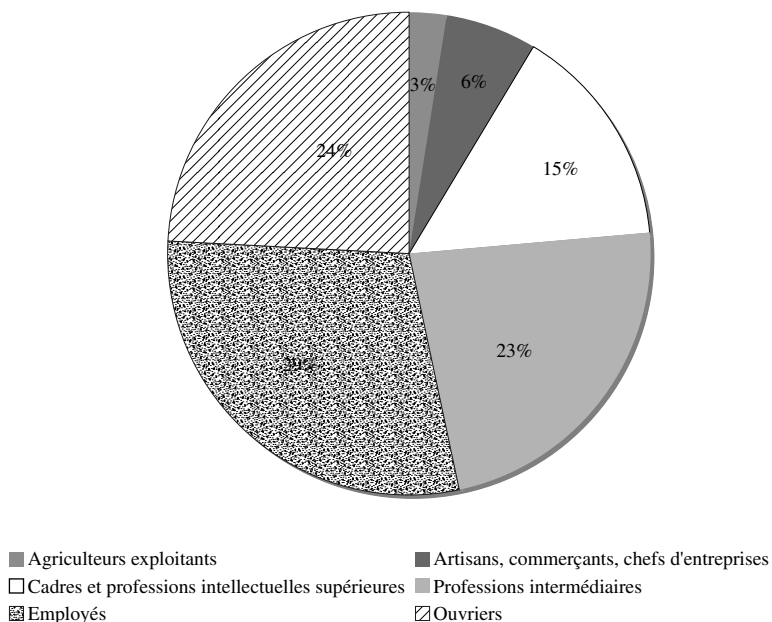
Catégories sociales en 1982



Catégories sociales en 1990

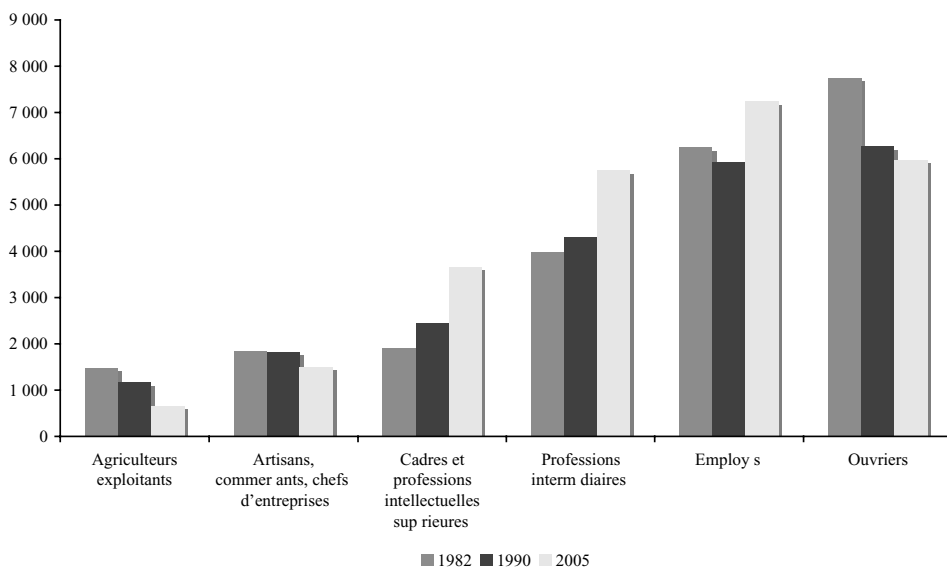


Catégories sociales en 2005



3. Nous utilisons un diagramme qui fait apparaître les évolutions des différentes catégories sociales pour les années étudiées.

Évolution des différentes catégories sociales



Ce graphique montre bien la diminution absolue du nombre des indépendants « Agriculteurs exploitants », « Artisans, commerçants et chefs d’entreprise », en particulier des agriculteurs, ainsi que de la catégorie « Ouvriers ». La catégorie sociale modale en France en 2005 est celle des employés, dont nous savons par ailleurs que le taux de féminisation est élevé. Ce graphique illustre également l’importance croissante de la catégorie des « Professions intermédiaires » presque aussi nombreuse que celle des « Ouvriers » ainsi que la place croissante des « Cadres ».

13. ÉVOLUTION DES REJETS DE DIOXYDE DE SOUFRE

Mots-clefs

Pourcentage, graphique en secteurs

Énoncé

L’évolution de la pollution en dioxyde de soufre en France est donnée dans le tableau suivant par secteur.

La structure des rejets par source

	1980 (%)	1990 (%)	1994 (%)
Résidences et bureaux	12,6	15,1	14,5
Industrie	31,8	21,6	22,0
Centrales électrothermiques	36,5	26,1	17,3
Transformations d’énergie	6,3	10,2	12,6
Procédés industriels	9,0	14,9	16,5
Transports	3,8	12,1	17,2
Ensemble	100,0	100,0	100,0

Source : TEF 93/94.

Cette pollution représentait 3 348 milliers de tonnes en 1980 et 1 200 milliers de tonnes en 1990, 961 milliers de tonnes en 1994.

Représentez graphiquement ces trois distributions ; le graphique devra rendre compte de la décroissance relative du phénomène.

Corrigé

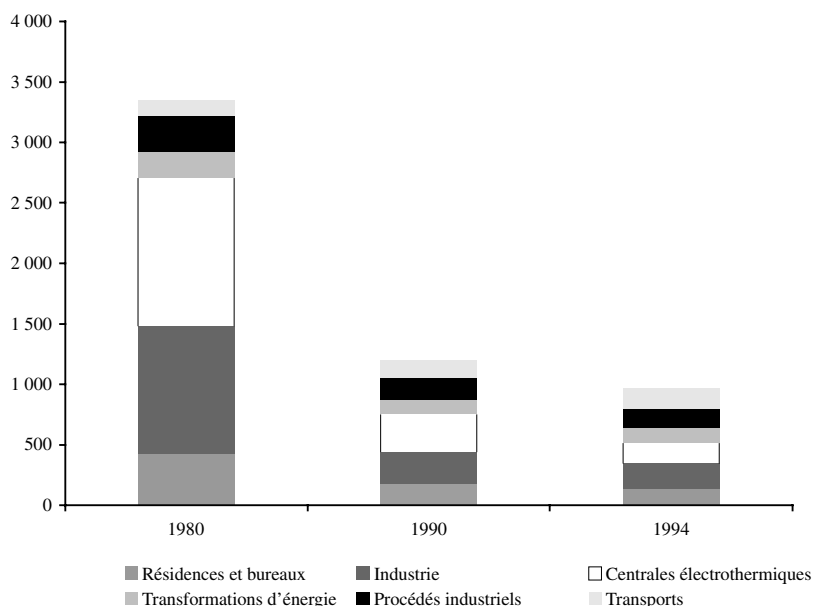
Parmi les possibilités de représentations graphiques, une solution est de faire apparaître d’une part les modifications relatives des sources de pollution au dioxyde, et d’autre part les évolutions de la pollution au cours du temps. Ce qui nous conduit à construire le tableau des quantités.

Les quantités de rejet

	1980	1990	1994
Résidences et bureaux	421,8	181,2	139,3
Industrie	1064,7	259,2	211,4
Centrales électrothermiques	1222,0	313,2	166,3
Transformations d'énergie	210,9	122,4	121,1
Procédés industriels	301,3	178,8	158,6
Transports	127,2	145,2	165,3
Ensemble	3348,0	1200,0	961,0

Le premier graphique retenu est en « tuyaux d'orgue », il montre la forte décroissance du volume de la pollution en dioxyde de soufre. Les modifications de l'importance relative des sources de pollution est peu lisible.

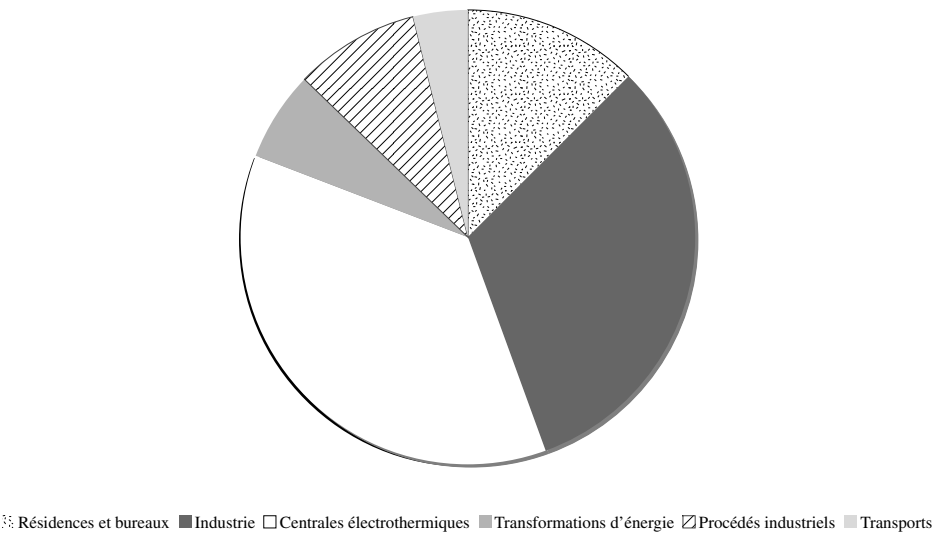
Évolution du volume des rejets



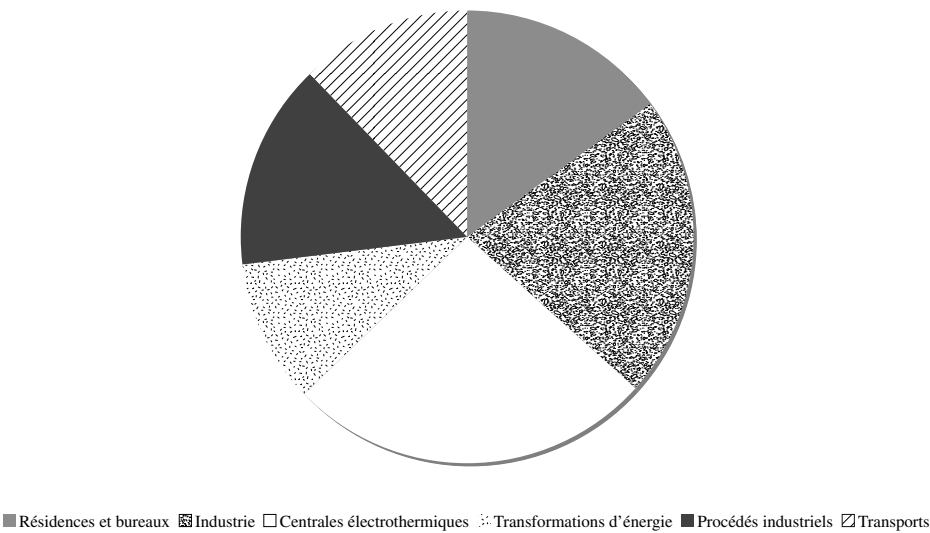
Le volume de la pollution s'est considérablement réduit entre 1980 et 1994. Parmi les raisons expliquant cette évolution, nous pouvons évoquer la lutte contre la pollution et la diminution des activités industrielles sur le territoire de la France.

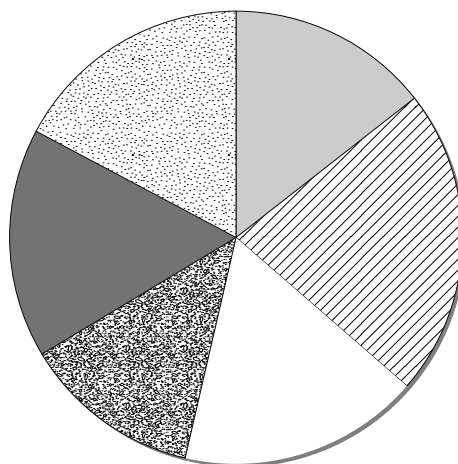
Pour mettre en lumière les modifications de la structure des diverses sources de rejets de dioxyde de soufre nous utiliserons des graphiques en secteurs.

Répartition des pollutions en 1980



Répartition des pollutions en 1990



Répartition des pollutions en 1994

■ Résidences et bureaux ▨ Industrie □ Centrales électrothermiques ▩ Transformations d'énergie ▤ Procédés industriels ▧ Transports

Cet ensemble de graphiques permet de voir que la réduction de la pollution n'est pas homothétique. Très globalement, les activités de consommation (transports, résidences) voient leur importance augmenter relativement aux rejets industriels. La réduction importante de la pollution se traduit par une répartition différente des principales sources de production.

14. HISTOGRAMME D'UN TABLEUR ET HISTOGRAMME STATISTIQUE

Mots-clefs

Histogramme, variable continue

Énoncé

Les tableurs disponibles proposent une représentation graphique baptisée « histogramme ». Cet exercice vise à expliciter la différence entre les deux représentations, il suppose que le lecteur dispose d'un tableur.

Soit la répartition des revenus d'une population présentée dans le tableau suivant en milliers d'euros.

1. Représentez cette distribution à l'aide de la fonction « histogramme » d'un tableur.
2. Construisez l'histogramme statistique de la série.

Répartition des revenus

Classes	Effectifs (n_i)
[0 ; 40[4 500
[40 ; 50[8 500
[50 ; 60[9 000
[60 ; 70[8 500
[70 ; 80[6 500
[80 ; 90[6 000
[90 ; 100[6 000
[100 ; 120[6 000
[120 ; 150[7 000
[150 ; 180[6 000
[180 ; 250[3 000
[300 ; 400[3 500
	74 500

Corrigé

1. Nous allons tout d’abord construire l’ « histogramme » avec cette fonction d’un tableur.

Données pour « l’histogramme » du tableur

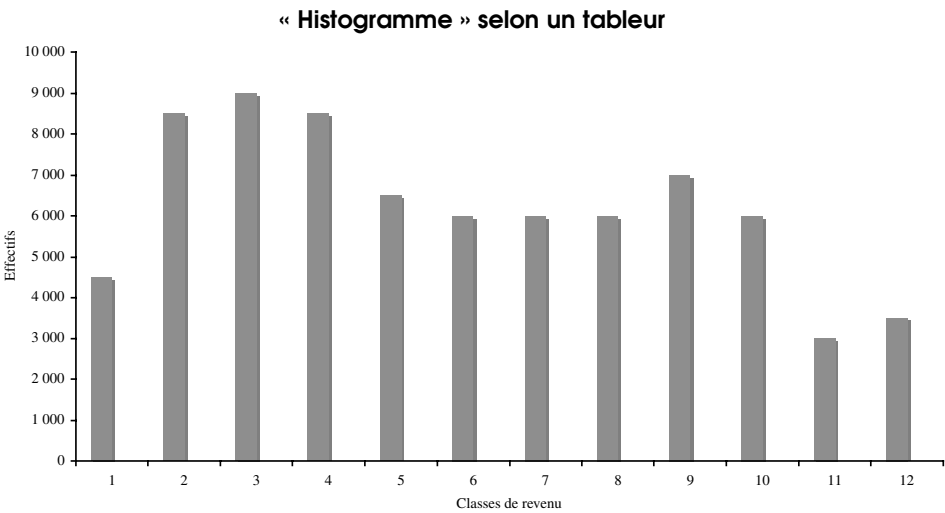
Classes	Effectifs
1	4 500
2	8 500
3	9 000
4	8 500
5	6 500
6	6 000
7	6 000
8	6 000
9	7 000
10	6 000
11	3 000
12	3 500

Nous obtenons le graphique ci-contre.

En quoi cette représentation est-elle insatisfaisante ?

La première raison tient au fait que la représentation donne une image d’une variable discrète, or nous avons une variable continue. La contradiction entre le caractère continu de la variable et sa représentation graphique sous forme discrète rend sans intérêt une telle « représentation » qui justement ne représente pas la nature de la variable.

La seconde raison du caractère insatisfaisant de cette représentation s’explique par l’absence de prise en compte des différences dans l’amplitude des classes.



Cette fois encore le graphique est infidèle à ce qu’il veut représenter.

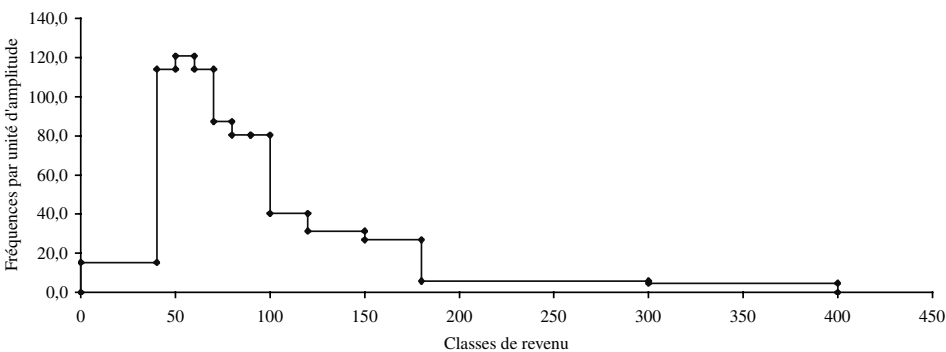
2. L’histogramme statistique va rendre compte du caractère continu de la variable et des amplitudes différentes.

Tableau statistique pour la construction de l’histogramme statistique

Classes	n_i	c_i	f_i	$\frac{f_i}{a_i} \times 100$
[0 ; 40[4 500	20,0	6,0	15,1
[40 ; 50[8 500	45,0	11,4	114,1
[50 ; 60[9 000	55,0	12,1	120,8
[60 ; 70[8 500	65,0	11,4	114,1
[70 ; 80[6 500	75,0	8,7	87,2
[80 ; 90[6 000	85,0	8,1	80,5
[90 ; 100[6 000	95,0	8,1	80,5
[100 ; 120[6 000	110,0	8,1	40,3
[120 ; 150[7 000	135,0	9,4	31,3
[150 ; 180[6 000	165,0	8,1	26,8
[180 ; 250[3 000	215,0	4,0	5,8
[300 ; 400[3 500	350,0	4,7	4,7
	74 500		100,0	

Le graphique suivant rend compte du caractère continu de la variable et du fait que les classes ont des amplitudes différentes. Cet histogramme est également construit à l’aide d un tableur mais pas en utilisant la fonction histogramme.

Histogramme de la distribution des revenus



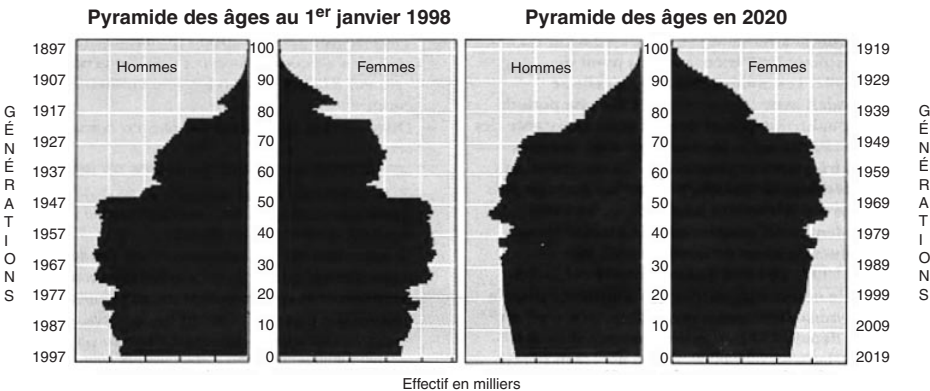
15. ANALYSE DE LA PYRAMIDE DES ÂGES

Mots-clefs

Histogramme, analyse de graphique, représentation graphique

Énoncé

Pyramides des âges de la population française en 1998 et 2020



Population par groupe d'âge Projection* de population par groupe d'âge
Âge moyen au 1^{er} janvier à l'horizon 2020

Années	Moins de 20 ans %	20 ans à 59 ans %	60 ans et plus %	Âge moyen années	Années	Moins de 20 ans %	20 ans à 59 ans %	60 ans et plus %	Population totale milliers
1946	29,5	54,5	16,0	35,6	2000	25,9	53,6	20,5	59 412
1970	33,2	48,8	18,0	34,8	2005	25,0	54,2	20,8	60 642
1980	30,6	52,4	17,0	35,7	2010	24,2	53,0	22,8	61 721
1990	27,8	53,2	19,0	36,9	2015	23,4	51,7	24,9	62 648
1996	26,0	54,0	20,0	37,9	2020	22,7	50,5	26,8	63 453
1997 <i>p</i>	25,9	53,8	20,3	38,1	* Hypothèse de fécondité = 1,8 enfant par femme.				
1998 <i>p</i>	25,8	51,8	20,4	38,3					

Source : TEF 98/99

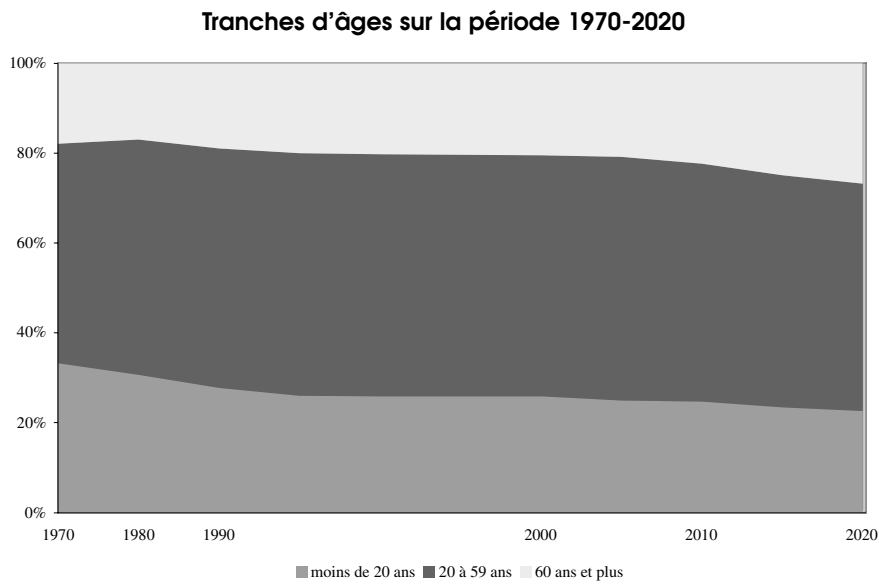
1. À quel type de représentation graphique fait appel la pyramide des âges ?
2. Quelles réflexions vous inspirent la comparaison entre les deux pyramides ?
3. Comparez la distribution des hommes à celles des femmes en 1998 ? Vos conclusions sont-elles les mêmes en 2020 ?
4. La population est-elle en croissance ? Justifiez votre réponse.
5. La population est-elle vieillissante ? Justifiez votre réponse.
6. Représentez l'évolution des tranches d'âge sur la période 1970/2020.

Corrigé

1. Les variables représentées sont les âges des hommes et des femmes, vivant à une date donnée. Il s'agit donc de variables continues. Les pyramides sont des histogrammes horizontaux.
2. La comparaison entre la pyramide des âges au 1^{er} janvier 1998 et celle qui est prévue en 2020 montre un épaississement de la courbe des plus de 50 ans accompagné de la disparition du creux de la génération 1915, le creux de la génération 1940 étant encore perceptible. Les jeunes générations sont beaucoup plus lisses dans la pyramide de 2020 car les chiffres utilisés sont des prévisions de population. On ne projette pas une augmentation du nombre de naissances en 2020.
3. En janvier 1998, on constate qu'il naît légèrement plus de garçons que de filles mais que, plus on avance en âge, et plus la pyramide des femmes s'épaissit par rapport à celle des hommes. Les creux des deux guerres sont aussi perceptibles chez les hommes que chez les femmes.

Ces tendances se maintiennent en 2020. On voit, toutefois, d'une façon plus nette l'épaississement de l'histogramme des femmes pour les tranches âgées de la population.

4. Le tableau présentant la projection de la population indique une croissance prévue de la population à l'horizon 2020.
5. Nous voyons sur la pyramide des âges que globalement la population est en croissance du fait des ses tranches vieillissantes. Ce qui est confirmé par les chiffres des tableaux. Nous constatons une proportion croissante des personnes de plus de 60 ans, une réduction de la part de la population jeune (moins de 20 ans). Ce vieillissement s'apprécie également par l'élévation continue de l'âge moyen de la population.
6. Le graphique ci-contre, en bandes superposées, illustre l'analyse numérique.



16. DISTRIBUTION DE SALARIÉS

Mots-clefs

Histogramme, polygone des fréquences

Énoncé

La répartition d’une population de salariés par établissement au sein d’un bassin d’emploi est donnée par le tableau ci-dessous.

- 1. Construisez l’histogramme.
- 2. Construisez le polygone des fréquences.

Répartition des salariés par établissement	
Classes	Effectifs
[0 ; 20[925
[20 ; 40[1 665
[40 ; 50[2 590
[50 ; 60[3 145
[60 ; 80[2 775
[80 ; 110[2 405
[110 ; 160[1 850
[160 ; 240[1 480
[240 ; 360[1 110
[360 ; 500]	555
	18 500

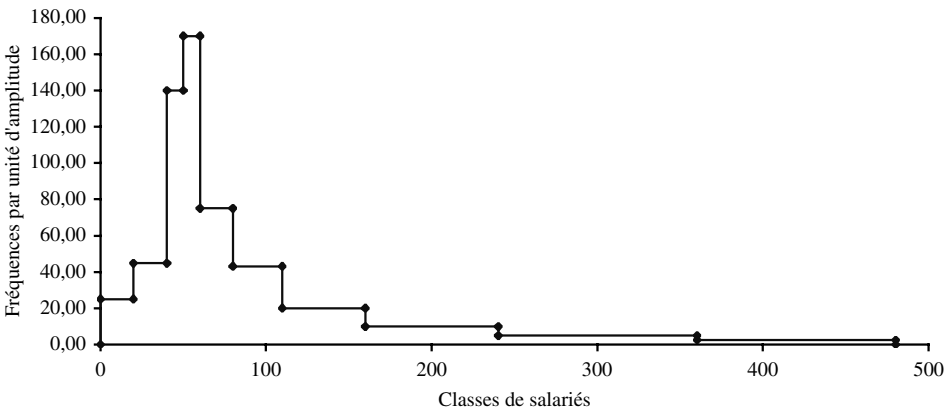
Corrigé

1. Pour construire l’histogramme, nous élaborons le tableau statistique nécessaire.

Calculs pour l’histogramme

Classes	n_i	a_i	c_i	f_i	$\frac{f_i}{a_i} \times 100$
[0 ; 20[925	20	10	5,0	25,0
[20 ; 40[1 665	20	30	9,0	45,0
[40 ; 50[2 590	10	45	14,0	140,0
[50 ; 60[3 145	10	55	17,0	170,0
[60 ; 80[2 775	20	70	15,0	75,0
[80 ; 110[2 405	30	95	13,0	43,3
[110 ; 160[1 850	50	135	10,0	20,0
[160 ; 240[1 480	80	200	8,0	10,0
[240 ; 360[1 110	120	300	6,0	5,0
[360 ; 500]	555	140	430	3,0	2,1
	18 500			100,0	

Histogramme de la distribution

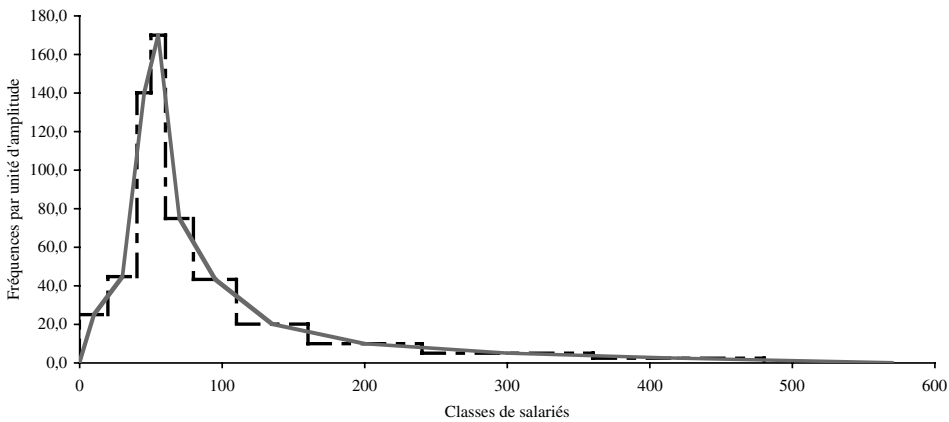


2. Nous construisons le polygone des fréquences en joignant les milieux de chaque sommet des rectangles. Les ordonnées sont les fréquences par unité d’amplitude, les abscisses sont des centres de classe si les classes sont d’amplitude égale. Pour fermer la courbe nous avons pris le point origine et un point qui se trouve à une distance égale à la moitié de l’amplitude de la dernière classe. Un tableau explicite les valeurs prises.

Valeurs pour le polygone des fréquences

c_i	$\frac{f_i}{a_i} \times 100$
0	0
10	25,0
30	45,0
45	140,0
55	170,0
70	75,0
95	43,3
135	20,0
200	10,0
300	5,0
430	2,1
570	0

Polygone des fréquences



Le polygone des fréquences donne une vision plus réaliste de la distribution en éliminant les ruptures entre les classes. Il permet également de percevoir la dissymétrie de la distribution.

Ce polygone de fréquences ne respecte pas le principe de la conservation des aires. Il n'est donc pas rigoureusement satisfaisant du point de vue statistique. Néanmoins, compte tenu de la diversité de l'amplitude des classes et de leur étendue, il permet de fournir une première image de la distribution des fréquences.

17. RÉPARTITION DES SUBVENTIONS AGRICOLES

Mots-clefs

Polygone des fréquences

Énoncé

Nous connaissons la valeur des subventions versées à une population d'agriculteurs.

Répartition des subventions par exploitation

Classes	Effectifs
[10 ; 20[12
[20 ; 30[24
[30 ; 40[36
[40 ; 50[20
[50 ; 70]	30
	122

1. Construisez l'histogramme.
2. Construisez le polygone des fréquences.

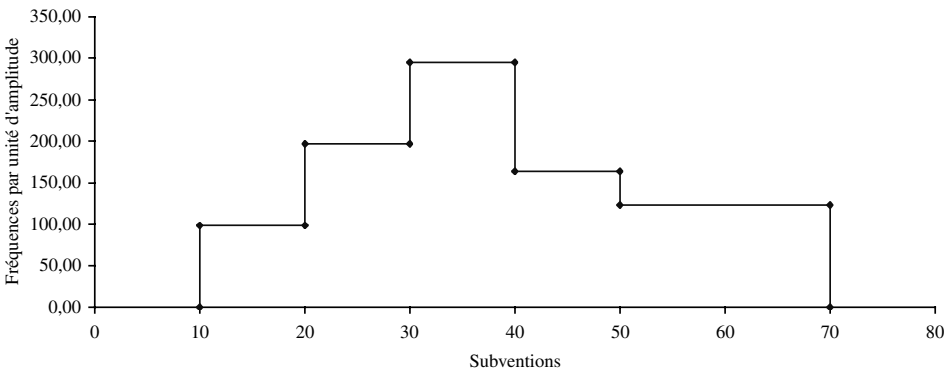
Corrigé

1. Pour construire l'histogramme, nous élaborons le tableau statistique nécessaire.

Calculs pour la construction de l'histogramme

Classes	b_i	b_{i+1}	a_i	c_i	n_i	f_i	$\frac{f_i}{a_i} \times 100$
[10 ; 20[10	20	10	15	12	9,84	98,4
[20 ; 30[20	30	10	25	24	19,67	196,7
[30 ; 40[30	40	10	35	36	29,51	295,1
[40 ; 50[40	50	10	45	20	16,39	163,9
[50 ; 70]	50	70	20	60	30	24,59	123,0
					122	100,00	

Histogramme de la distribution

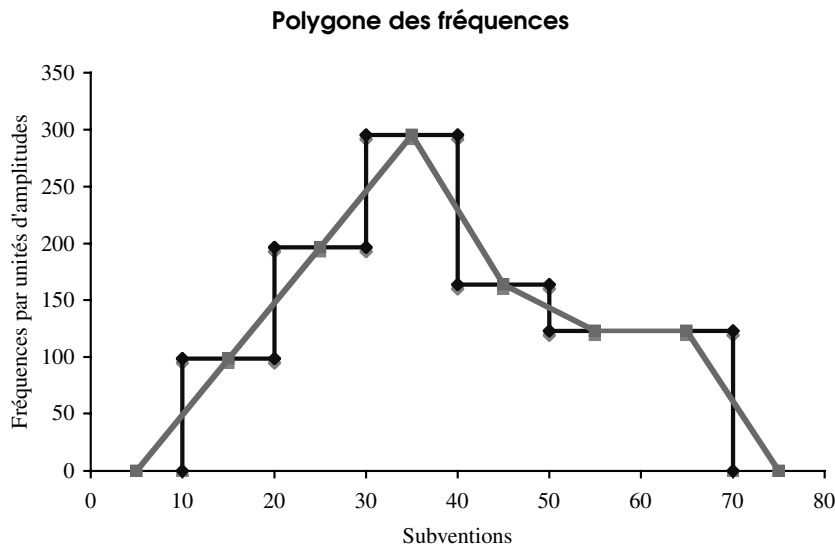


2. Le polygone des fréquences représente la distribution des fréquences. Nous le construisons en joignant les points de chaque milieu des segments de l’histogramme. Les ordonnées sont les fréquences par unité d’amplitude, les abscisses sont des centres de classe si les classes sont d’amplitude égale. Pour fermer la courbe nous avons pris le point origine et un point qui se trouve à une distance égale à la moitié de l’amplitude de la dernière classe. Un tableau explicite les valeurs prises.

Fréquences par unité d’amplitude

c_i	$\frac{f_i}{a_i} \times 100$
5	0
15	98,4
25	196,7
35	295,1
45	163,9
55	123,0
65	123,0
75	0

L’unité d’amplitude est de 10. Pour la dernière classe dont l’amplitude est de deux unités, nous n’avons pas retenu le centre de classe, nous avons subdivisé cette classe en classes ayant l’amplitude unitaire, ce qui nous conduit à créer deux classes. Nous considérons qu’entre 65 et 75, la répartition est stable. L’objectif est de faire en sorte que la surface du polygone des fréquences soit identique à celle de l’histogramme.



Le polygone des fréquences donne une vision plus réaliste de la distribution en éliminant les ruptures entre les classes. Il permet également de percevoir la dissymétrie de la distribution.

18. REPRÉSENTER DES INDICATEURS ENTRE MEMBRES DE L’UE

Mots-clefs

Diagramme polaire

Énoncé

Nous disposons pour quelques pays de l’Union européenne de trois indicateurs importants, ce sont des données relatives au quatrième trimestre 2005.

Les indicateurs

	Croissance annuelle du PIB en volume	Hausse annuelle des prix à la consommation	<u>Dette publique</u> PIB
Allemagne	1,6	2,1	66,4
France	1,2	2,0	65,1
Royaume-Uni	1,8	1,9	41,5
Espagne	3,5	4,1	46,9
Italie	0,1	2,2	106,5
Estonie	10,4	4,5	5,5

Réalisez une représentation graphique de ces séries permettant de visualiser le profil de chacun des pays relativement aux trois indicateurs.

Corrigé

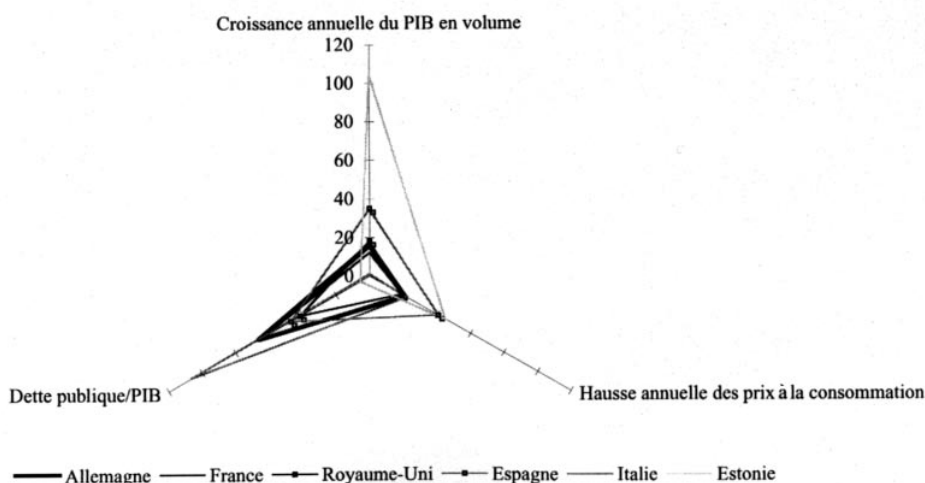
Il apparaît logique de vouloir représenter cette distribution avec un graphique ayant trois axes. Si nous représentons directement les données, nous n’aurons pas un graphique très satisfaisant puisque les valeurs prises par la part de la dette publique dans le PIB sont d’un ordre de grandeur différent des deux autres variables. Pour rendre les valeurs plus comparables nous avons multiplié par 10, pour les besoins du graphique, les valeurs du PIB en volume et des indices de prix à la consommation. Nous obtenons alors le tableau pour la représentation graphique.

Valeurs initiales

	Croissance annuelle du PIB en volume	Hausse annuelle des prix à la consommation	<u>Dette publique</u> PIB
Allemagne	1,6	2,1	66,4
France	1,2	2,0	65,1
Royaume-Uni	1,8	1,9	41,5
Espagne	3,5	4,1	46,9
Italie	0,1	2,2	106,5
Estonie	10,4	4,5	5,5

Valeurs utilisées pour la représentation graphique

	Croissance annuelle du PIB en volume	Hausse annuelle des prix à la consommation	<u>Dette publique</u> PIB
Allemagne	16	21	66,4
France	12	20	65,1
Royaume-Uni	18	19	41,5
Espagne	35	41	46,9
Italie	1	22	106,5
Estonie	104	45	5,5



Ce tableau montre l'opposition entre les pays ayant une dette publique forte (Italie, France, Allemagne) au pays ayant une dette moyenne (Royaume-Uni, Espagne) à un pays très peu endetté l'Estonie, qui est également un nouvel adhérent de l'UE. L'Estonie se caractérise également par un taux de croissance du pays, PIB en volume, très supérieur aux autres pays de l'échantillon.

19. RÉPARTITION D'UNE POPULATION SELON LES PCS ET LE GENRE

Mots-clefs

Double distribution, graphique en barres

Énoncé

Le tableau suivant donne la répartition de la population active par catégorie socioprofessionnelle (C.S.P.) des individus et par sexe pour une population donnée en 1989.

Les catégories socioprofessionnelles en 1989 (en milliers)

Catégories socioprofessionnelles	Hommes	Femmes
Agriculteurs exploitants	799	469
Artisans commerçants et chefs d'entreprises	1 160	597
Cadres et professions intellectuelles supérieures	1 644	671
Professions intermédiaires	2 618	1 975
Employés	1 625	5 146
Ouvriers	5 626	1 495

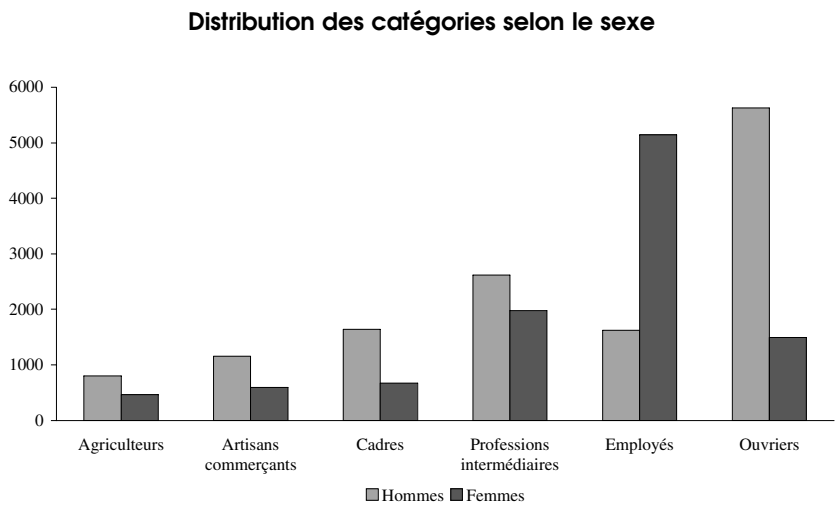
Source : INSEE

- 1. Représentez graphiquement cette double distribution.
- 2. Représentez les distributions conditionnelles selon diverses possibilités.

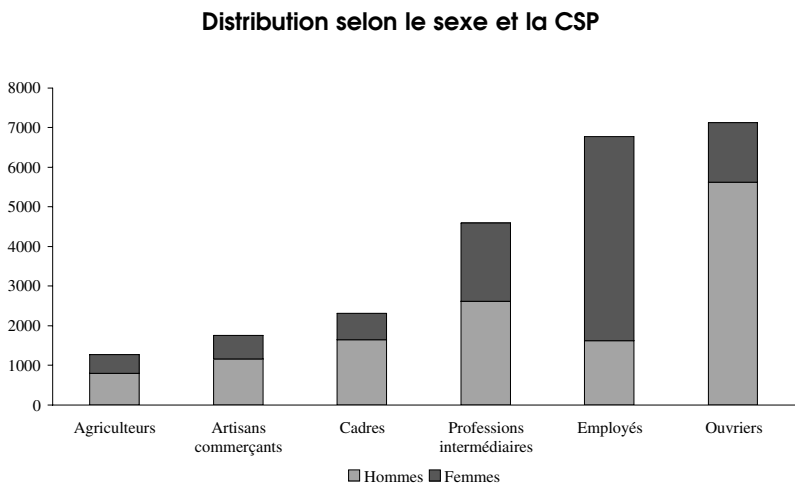
Corrigé

La représentation en tuyaux d’orgue est dans le cas de variables qualitatives la meilleure représentation graphique. Plusieurs solutions sont possibles.

- 1. Nous représentons côte à côte les deux distributions en tenant compte des effectifs de chaque catégorie.



- 2. Dans la deuxième représentation, nous utilisons des cartouches superposées rendant mieux compte de l’importance relative des actifs masculins et féminins, en gardant une représentation des valeurs absolues.



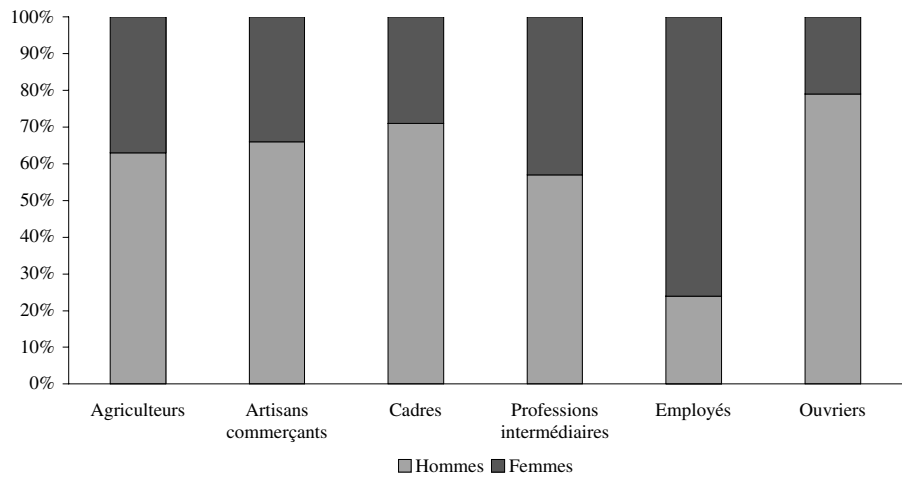
Par rapport au graphique précédent, nous constatons, entre autres, qu’il y a plus d’ouvriers que d’employés en 1989 sans distinction de genre.

3. Nous utilisons un graphique superposé pour représenter les distributions conditionnelles suivant la catégorie sociale et le genre dans un diagramme à cumul interne.

Tableau des fréquences relatives

Catégories	Hommes	f_{ih}	Femmes	f_{if}	Ensemble	f_i
Agriculteurs	799	0,63	469	0,37	1 268	1,00
Artisans	1 160	0,66	597	0,34	1 757	1,00
Cadres	1 644	0,71	671	0,29	2 315	1,00
Professions intermédiaires	2 618	0,57	1 975	0,43	4 593	1,00
Employés	1 625	0,24	5 146	0,76	6 771	1,00
Ouvriers	5 626	0,79	1 495	0,21	7 121	1,00

Part des femmes dans chacune des CSP



Cette représentation rend sensible le fait que la très grande majorité des employés sont des employées, alors que dans les autres catégories les hommes sont majoritaires.

20. RÉPARTITION DES ENTREPRISES PAR SECTEUR D'ACTIVITÉ
ET NOMBRE DE SALARIÉS

Mots-clefs

Représentation graphique en double barres

Énoncé

Répartition des entreprises par nombre de salariés et activités au 1/1/1998
(en milliers)

	Petites entreprises		Moyennes entreprises				Grandes entreprises	Total
	0	1 à 9	10 à 49	50 à 199	200 à 499	Total	≥ 500	
IAA	17,47	44,59	5,63	1,09	0,24	6,96	0,12	69,13
Industrie hors IAA	65,95	81,71	29,55	6,97	1,55	38,07	0,82	186,55
Construction	126,02	149,77	18,61	1,77	0,25	20,62	0,11	296,52
Commerce	275,31	293,14	34,86	4,44	0,61	39,92	0,28	608,64
Transports	47,10	28,55	7,52	1,51	0,29	9,32	0,10	85,07
Hôtellerie	78,65	109,38	8,05	0,59	0,07	8,72	0,06	196,81
Services aux entreprises	214,73	145,39	19,34	2,78	0,62	22,75	0,33	383,20
Services aux ménages	288,86	141,93	7,61	1,19	0,19	8,98	0,04	439,81
Total	1114,08	994,45	131,18	20,33	3,83	155,34	1,86	2265,72

1. Quelles sont les variables étudiées ? Quel est leur type ?
2. Proposez une représentation graphique permettant de visualiser la répartition des activités au sein de chaque groupe de moyennes entreprises.

Corrigé

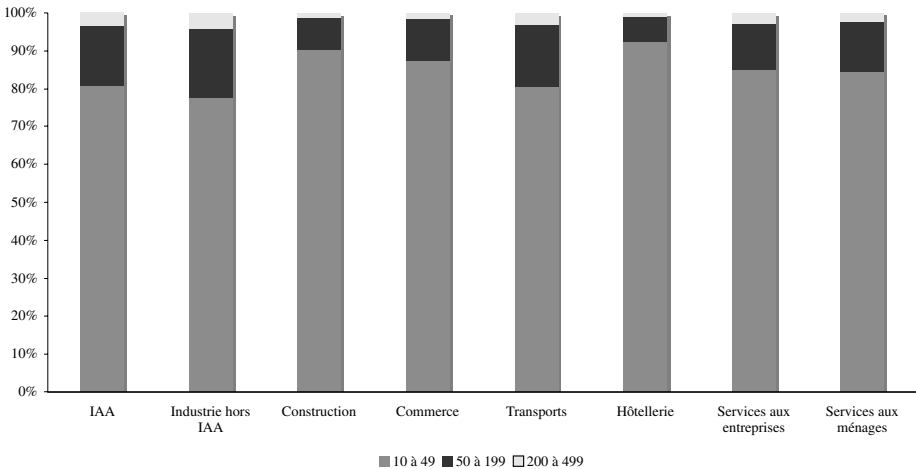
1. Nous avons un tableau croisant une variable qualitative (une nomenclature des activités) et une variable quantitative continue (le nombre de salariés). Cette variable est continue car, en équivalent temps plein, toutes les valeurs de la variable peuvent être atteintes.

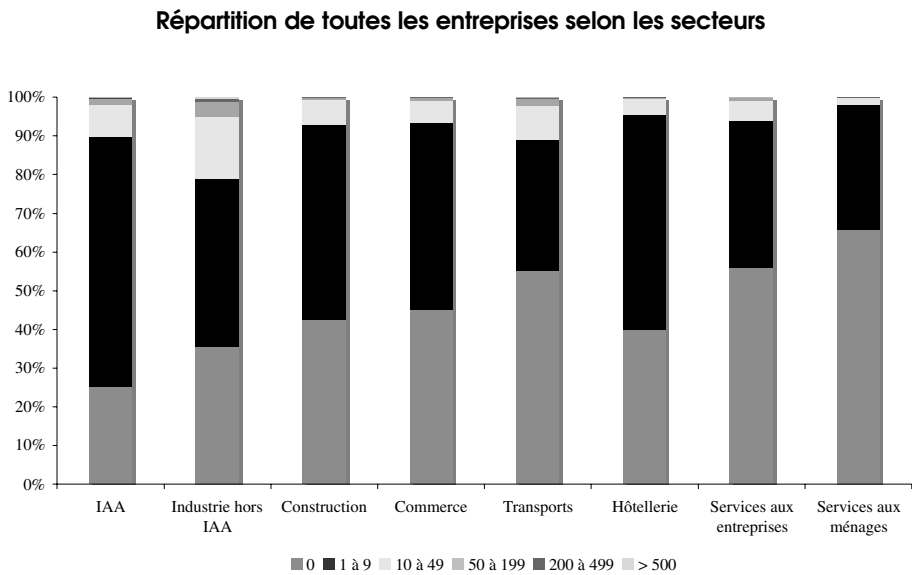
2.

Tableau de la répartition des entreprises pour construire le graphique (en %)

	Petites entreprises		Moyennes entreprises			Grandes entreprises	Total
	0	1 à 9	10 à 49	50 à 199	200 à 499	≥ 500	
IAA	25,3	65,4	8,1	1,6	0,3	0,2	100,0
Industrie hors IAA	35,4	43,8	15,8	3,7	0,8	0,4	100,0
Construction	42,5	50,5	6,3	0,6	0,1	0,0	100,0
Commerce	45,2	48,2	5,7	0,7	0,1	0,0	100,0
Transports	55,4	33,6	8,8	1,8	0,3	0,1	100,0
Hôtellerie	40,0	55,6	4,1	0,3	0,0	0,0	100,0
Services aux entreprises	56,0	37,9	5,0	0,7	0,2	0,1	100,0
Services aux ménages	65,7	32,3	1,7	0,3	0,0	0,0	100,0
Total	49,2	43,9	5,8	0,9	0,2	0,1	100,0

Répartition des moyennes entreprises par taille en fonction du secteur d'activités





Il est clair que les très grandes entreprises n'apparaissent pas sur le graphique.

21. GRAPHIQUE DES RECETTES ET DES DÉPENSES

Mots-clefs

Double graphique

Énoncé

Soit les éléments de l'exécution du budget d'une université grenobloise.

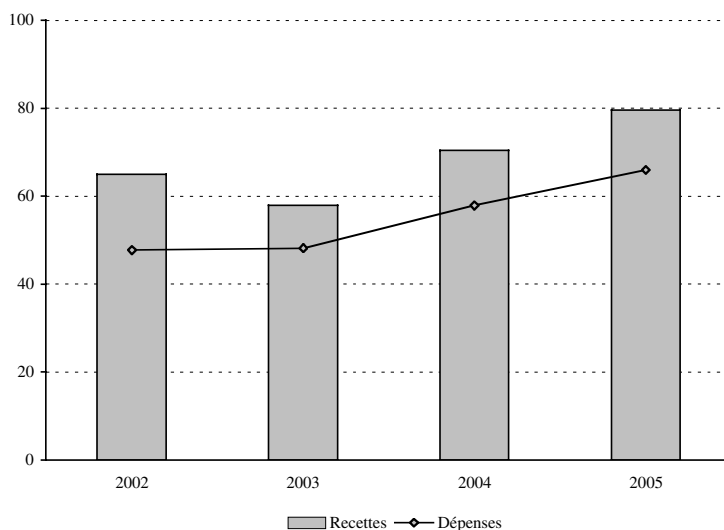
	2002	2003	2004	2005
Recettes	65,0	58,0	70,5	79,6
Dépenses	47,7	48,2	57,9	66,0

Donnez une représentation graphique de ces deux distributions.

Corrigé

Pour représenter cette double distribution nous utilisons deux formes de représentation : les barres pour les recettes, et une courbe pour les dépenses.

Les taux d'exécution budgétaires



22. SOUTENANCE DE THÈSES

Mots-clefs

Pourcentage, multiplicateur

Énoncé

L'évolution des soutenances de thèse dans une université de Grenoble a évolué comme ci-dessous.

	1999	2000	2001	2002
Effectifs	425	448	431	406
% de femmes	34,5	31,5	30	25,5
% de post-doctorants	20	36,5	38,5	36

Source : publication de l'UJF

1. Calculez le taux d'évolution du nombre de thèses entre 1999 et 2002.
2. Le pourcentage des femmes décroît-il plus rapidement que celui des effectifs ?
3. Peut-on constater une réduction du nombre de thèses soutenues par les hommes ?

Corrigé

1. Nous constatons sur la période une réduction du nombre de thèses, pour quantifier cette diminution, nous calculons les multiplicateurs selon la formule usuelle :

$$(1 + r_1) = \frac{X_1}{X_0}$$

Par exemple entre 2000 et 1999 le multiplicateur est : $1 + r = \frac{448}{425} = 1,054$.
L'habitude est de multiplier le résultat par 100 pour faciliter la lecture. Dans les calculs, il est nécessaire de ne pas oublier cette convention.

Multiplicateurs des thèses

	2000/1999	2001/2000	2002/2001
Multiplicateurs annuels	105,4	96,2	94,2

Entre 1999 et 2001 le multiplicateur est de :

$$(1 + r) = 1,054 \cdot 0,962 \cdot 0,942 \cong 0,955$$

Le multiplicateur entre 1999 et 2002 est de 95,5. Le taux de réduction sur la période est de 4,5 %.

Nous pouvons également le calculer directement $(1 + r) = \frac{406}{425} = 0,955$.

2. Pour obtenir le taux de réduction du nombre de thèses soutenues par les femmes, nous calculons le nombre de thèses soutenues par les femmes puis le multiplicateur entre 1999 et 2002. Il est aussi possible d'utiliser les calculs précédents. Nous allons présenter les deux calculs.

Calcul du nombre de thèses soutenues par des femmes :

Nombre de thèses \times pourcentage de thèse soutenues par des femmes

Par exemple pour 2000 : $448 \times 0,315 = 141,12$ soit 141 thèses.

Évolution du nombre de thèses soutenues par des femmes

Années	1999	2000	2001	2002	Produit des multiplicateurs annuels
Nombre de thèses soutenues	147	141	129	104	
Multiplicateurs		96,24	91,62	80,1	70,60

Soit $(1 + r) = 0,9624 \cdot 0,9162 \cdot 0,801 \cong 0,706$,

Le multiplicateur est de 70,6.

Le multiplicateur global est de 95,5 ; pour les femmes, il est de 70,6. Le taux de réduction de thèses « féminines » atteint 29,6 %.

Nous pouvons obtenir le même résultat sans avoir à calculer le nombre de thèses soutenues par les femmes.

Soit n_i le nombre total de thèses soutenues l'année i , n'_i le nombre de thèses soutenues par des femmes et f_i la proportion de ces thèses, $n'_i = f_i \cdot n_i$.

Le multiplicateur pour les thèses « féminines » entre $i - 1$ et i est :

$$\frac{n'_i}{n'_{i-1}} = \frac{f_i \cdot n_i}{f_{i-1} \cdot n_{i-1}} = \frac{f_i}{f_{i-1}} \cdot (1 + r)_i$$

Exemple numérique entre 2001 et 2002, le multiplicateur global est 94,2 ; le multiplicateur pour les femmes est de $94,2 \cdot \frac{25,5}{30} = 80,07$.

Nous pouvons également le calculer directement $(1 + r)_f = \frac{406}{425} \cdot \frac{25,5}{34,5} = 0,706$. Le multiplicateur est de 70,6.

3. Nous calculons directement le multiplicateur des thèses « masculines ».

$$(1 + r)_h = \frac{406}{425} \cdot \frac{(100 - 25,5)}{(100 - 34,5)} = 1,0866$$

Le multiplicateur est de 108,7.

Le nombre de thèses soutenues par des hommes est plus important en 2002 qu'en 1999, il a augmenté de plus de 8 %. La réduction du nombre de thèses soutenues s'explique donc par la seule réduction du nombre des thèses « féminines ».

La diminution de 4,5 % du nombre de thèses soutenues s'explique par une réduction de 29,4 % des thèses « féminines » et une augmentation de 8,7 % des thèses « masculines ».

23. TAUX DE SURVIE DES ENTREPRISES

Mots-clefs

Pourcentage

Énoncé

Les créations d’entreprises en 1995

Ensemble	295 416
Soit en pourcentages, suivant la taille	
0 salarié	74,5
1 ou 2 salariés	17,1
3 à 5 salariés	5,5
6 à 9 salariés	1,5
10 salariés et plus	1,4

Taux de survie à trois et cinq ans des entreprises créées en 1995 (en %)

	Survie à trois ans	Survie à cinq ans
0 salarié	67	51
1 ou 2 salariés	75	61
3 à 5 salariés	83	70
6 à 9 salariés	85	74
10 salariés et plus	87	77

- 1. Définissez le taux de survie.
- 2. Calculez le taux de survie à trois ans des entreprises de moins de 5 salariés.
- 3. Déterminez le taux de survie à cinq ans des entreprises de plus de 6 salariés.

Corrigé

- 1. La durée de vie des entreprises est obtenue par la différence du nombre d’entreprises entre la date de cessation économique et celle de la création. Le taux de survie à 1, 3, 5 ou 7 ans est donc égal au nombre des entreprises de la génération N encore actives lors de leur 12^e, 36^e, 60^e ou 84^e mois, divisé par le nombre d’entreprises créées l’année N.
- 2. Pour calculer le taux de survie à trois ans des entreprises de moins de 5 salariés nous devons construire les tableaux statistiques suivants.

Les créations d'entreprises en 1995

Classe d'entreprises	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	n_i	f_i	p_i	F_i
0 salarié	220 085	0,745	74,5	0,745
1 ou 2 salariés	50 516	0,171	17,1	0,916
3 à 5 salariés	16 248	0,055	5,5	0,971
6 à 9 salariés	4 431	0,015	1,5	0,986
10 salariés et plus	4 136	0,014	1,4	1,000
	295 416	1,000	100,0	

La survie à trois ans

Classe d'entreprises	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	n_i	f_i	p_i	F_i
0 salarié	147 457	0,718	71,8	0,718
1 ou 2 salariés	37 887	0,184	18,4	0,903
3 à 5 salariés	12 656	0,062	6,2	0,964
6 à 9 salariés	3 766	0,018	1,8	0,982
10 salariés et plus	3 598	0,018	1,8	1,000
	205 364	1,000	100,0	

La survie à cinq ans

Classe d'entreprises	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	n_i	f_i	p_i	F_i
0 salarié	112 243	0,697	69,7	0,698
1 ou 2 salariés	30 815	0,192	19,2	0,889
3 à 5 salariés	11 374	0,071	7,1	0,960
6 à 9 salariés	3 279	0,020	2,0	0,980
10 salariés et plus	3 185	0,020	2,0	1,000
	160 896	1,000	100,0	

Le taux de survie à trois ans des entreprises d'au moins de 5 salariés est le rapport entre le nombre d'entreprise de 5 salariés au moins existant trois ans après leur création et le nombre d'entreprise créées en 1995.

Les entreprises existant après trois ans sont de 147 457 avec 0 salariés, 37 887 avec 1 ou 2 salariés, 12 656 de 3 à 5 salariés soit 198 000. Les entreprises créées étaient pour les mêmes catégories 220 805 avec 0 salariés, 50 516 avec 1 ou 2 salariés, 16 248 de 3 à 5 salariés soit 286 849.

Le taux de survie à trois ans est donc $\frac{198\,000}{220\,805} \cdot 100 = 69,03 \%$.

3. À partir du tableau de la survie des entreprises à cinq ans et celui des créations d'entreprise, nous pouvons calculer le taux de survie à cinq ans des entreprises de plus de 6 salariés comme le rapport du nombre d'entreprises de plus de 6 salariés existant cinq ans après leur création et le nombre d'entreprises de plus de 6 salariés créées en 1995.

Le nombre d'entreprises de six salariés cinq ans après leur création est égal au nombre total des entreprises multiplié par le pourcentage des entreprises de plus de 6 salariés : $160\,896 \cdot (1 - 0,96)$.

Le nombre d'entreprises de plus de six salariés créées en 1995 : $295\,416 \cdot (1 - 0,971)$.

Le taux cherché est donc $\frac{160\,896 \cdot (1 - 0,96)}{295\,416 \cdot (1 - 0,971)} \cdot 100 = 75,12 \%$